JOURNAL ARTICLE

Perceptual Associations Between Words and Speaker Age

Jonny Kim
University of Hawai'i at Mānoa, USA
jonnykim@hawaii.edu

Recognition proficiency was measured in a lexical decision experiment using Korean words including ones that are stereotypically associated with younger or older people. Both accuracy rates and response times were significantly improved when a word was spoken in a voice whose age matched the age category of the word based on stereotype-based age rating. Accuracy rates were also significantly improved when a word was spoken in a voice whose age matched the frequency-based age category, although the effect on response times was marginal. The results are discussed in the context of an exemplar-based account incorporating social exemplar activation. First, the activation of social information is influenced by distributional properties of lexical use across social groups, indicating that social information of speakers is not only linked to sub-lexical representations as shown in previous studies, but also directly linked to lexical representations through social indices. Furthermore, the presence of an additional effect of stereotypes expands our understanding of the speech perception mechanism, suggesting that the social indices are enhanced by stereotypical associations between words and age categories. When listening to words with a wide spectrum of age-related variability, word-stereotypes activate social information, guiding to the lexical representations via social indices.

Keywords: lexical recognition; word-stereotype; word frequency; speaker information; perceptual association; social exemplar activation

1 Introduction

During the past few decades, numerous words have been created and spread out by younger speakers of Korean, yielding a great deal of generational gap in terms of what lexical items are preferably used. This study explores the cognitive mechanism that governs the association between Korean words in the cross-generational variation and speakers from different social categories. Specifically, if a word is produced frequently and exclusively by a certain age group, might the word be recognized more easily when spoken by a speaker who is perceived to belong to the congruent age group than when spoken by a speaker who is not? If so, how is social information (e.g., speaker age) utilized during lexical access? How does it interact with the social characteristics of words, and how might we model the speech processing mechanism that is responsible for that interaction?

We examine two different types of perceptual associations. On the one hand, the associations are governed by listeners' experience, or the statistical distribution of their exposure to word occurrences over different age groups. Walker and Hay's (2011) lexical decision test found improved recognition when listeners' perceived age of the speaker matched the age of the social group who uses the given word most frequently. The associations between a word and age (*word age*, henceforth) were determined based on relative frequencies of word occurrence across two different New Zealand English (NZE) spoken

corpora recorded in different times. Words that appear relatively frequently in one corpus compared to the other were selected as *young words* and *old words,* respectively. Therefore, lexical representations are indexed to the words' distributional properties as determined by relative frequency of occurrences across age groups.

On the other hand, we also examine the possibility that the distributional associations are either enhanced or attenuated by the existence of social and/or linguistic stereotypes.[1] If so, word-stereotypes may have an additional effect on word recognition, which is not predicted solely by distributional properties. For example, among Walker and Hay's (2011) stimuli, the word *internet* is one of the young words, i.e., words that are spoken more frequently by younger NZE speakers, and thus yielded facilitated recognition when spoken by younger speakers in their experiment. However, its stereotypical association with the younger generation may not be as strong as, for example, that of *lol* (an internet-slang acronym, or more precisely an initialism, for *laughing out loud*). Additionally, the distributional association of *lol,* whether pronounced as [lɒl] or [ɛlɔʊɛl], may not be as strong as its own stereotypical association with the younger generation, because it is not as frequently spoken as it is used orthographically, lacking phonetic exemplars. To the best of our knowledge, no exemplar-based accounts for spoken word perception incorporate stereotypical associations as a component of lexical recognition, independent from distributional properties. We propose that statistical distribution across social groups plays a crucial role in the formation of social indices over stored exemplars, but the distributional association is reinforced by word-stereotypes, which are also indexed to the lexical representation, and thus affect lexical access.

2 Background

Although focusing less on word recognition than on the segmental level, sociophonetic research has examined how and to what degree social information attributed to individual speakers wields influence on speech perception. Johnson et al. (1999), for example, found that listeners' perceptual boundaries of resynthesized vowels between [ʊ] and [ʌ] were significantly affected by perceived gender of the voice and of a visually presented face in a video clip. In a follow-up experiment, participants were asked to imagine the talker as either male or female and listened to a gender-neutral voice without visual stimuli. Even though no concrete talker information was given, the perceptual boundaries of the participant group that was asked to imagine a female speaker were influenced mirroring the physiological difference of vocal tract length. Based on these results, Johnson et. al. conclude that the integration of visual and auditory sociolinguistic cues either in a concrete form (by visual cues) or in an abstract form (by imagination) take place in an early stage of speaker normalization. They also note that estimation of vocal tract size (e.g., Ladefoged & Broadbent, 1957) is not a consistently predictable parameter in reality, so that what listeners actually perceive are not only the concrete cues based on algorithmic sensory information concerning physical properties but also subjective and abstract representations of *talkers*[2] based on experiences and expectations.

---

[1] As opposed to *indicators* and *markers*, linguistic stereotypes refer to ones that exist at the conscious level of listeners' attention to stylistic variation, ones for which individuals in the speech community can discuss the relation between the linguistic variable and its associated social group (Labov, 1972). We follow this definition in this study.

[2] Many researchers in sociophonetics, including Johnson et al., differentiate 'speakers' and 'talkers,' implying that perceived talker identity is not necessarily identical with the speaker's identity. The latter means listener's "perceptual impression of a person's identity based on hearing or watching the person speak," which is a combination of "objective facts about a person, such as vocal tract length" (1999, p. 365) and socio-cultural factors.

Stemming from this viewpoint, the body of work demonstrating the effects of social information has been growing steadily over the past two decades. It has been shown that listeners can exploit or be biased by various pieces of social information about the speaker and their association with phonetic variants while resolving auditory ambiguity. The utilized speaker information includes age (Hay et al., 2006b; Koops et al., 2008; Drager, 2011), gender (Strand & Johnson, 1996; Johnson et al., 1999), regional dialect (Niedzielski, 1999), ethnicity (Staum-Casasanto, 2008), and social class (Hay et al., 2006b).

As for the effect of speaker age, in Drager's (2011) vowel identification test, NZE listeners heard resynthesized vowel tokens in a continuum between the TRAP and DRESS[3] vowels in various voices and identified which word they heard. The two vowels are involved in a chain shift in NZE, in which younger speakers' pronunciation of TRAP is raised to overlap with the space of DRESS. Each voice was paired with a photograph of either a younger or an older face. The older participant group (but not the younger group) tended to identify the ambiguous vowels as TRAP when younger faces were presented, showing that the perceptual boundaries of older listeners are biased based on the association of younger speakers with a raised vowel space of TRAP.

As mentioned in Section 1 Walker and Hay (2011) specifically provided evidence for the role of distributional associations in lexical-level processing. They argue that lexical recognition is affected by how much the incoming signal resembles the general phonetic properties of the social group who produces the word frequently, and lexical representation is shaped by a lifetime of exposure to the statistical distribution of phonetic realizations of different social groups. They strengthened the argument in a post hoc analysis, testing whether there was an effect of listeners' conscious awareness of the frequencies for their lexical stimuli. Conscious awareness of word frequencies was measured in a survey with a different group of participants, where participants rated, on a five-point scale, whether they think younger or older people are more likely to use each of the words. The results revealed no significant interaction between the rating and the voice each word was spoken in. That is, words with age-stereotypes were not recognized faster when spoken by an age-matched speaker. Note that what the post hoc survey measured was respondents' conscious metalinguistic judgments on words' association with age categories, and these measurements are consistent with the concept of *linguistic stereotype* in Labov's (1972) demarcation of linguistic variables (see note 1). Walker and Hay concluded that the processing advantage they observed was "driven by experience and not by a conscious awareness of word frequencies/social associations" (2011, p. 228).

Although little has been known about the effect of stereotypical associations on lexical processing, there is evidence (reviewed in Drager & Kirtley, to appear) to suggest that awareness of variation strengthens the effect of social information on segment perception (Niedzielski, 1999; Hay et al., 2006a; Hay & Drager, 2010). In Niedzielski's (1999) perception test, participants living in Detroit heard sentences containing the diphthong /aw/, naturally produced by a Detroiter, and chose the best matching vowel token from a set of resynthesized tokens on a continuum between /aw/ and its raised form. Participants who were told that the speaker was from Canada tended to choose the raised tokens as the best match, while participants who were told that the speaker was from Michigan did not. The results indicate that the Detroiters' perception is biased by a stereotype that Canadians produce a raised variant of /aw/. What is notable is that Detroit residents associate only Canadian speakers with the raised variant, when, in fact, the dialect spoken in Detroit also raises the diphthong.

---

[3] The lexical sets established by Wells (1982) are used here to refer to the vowel classes.

In addition, Hay and colleagues (Hay et al., 2006a; Hay & Drager, 2010) show that mere exposure to a social concept involved in a stereotype is sufficient to affect vowel perception. In both studies, participants matched naturally produced vowel tokens to vowels from a synthesized continuum between Australian versions and New Zealand versions. Prior to the tasks, participants were exposed to written labels of dialectal regions (Australia or New Zealand) at the top of the response sheet in Hay et al. (2006a), and stuffed-toy kangaroos and koalas (associated with Australia) or toy kiwis (associated with New Zealand) in Hay and Drager (2010). No explicit instructions were provided to the participants about how the materials were related to the task or to the speaker, but perception was biased towards the vowel tokens that are associated with the exposed concepts. Hay et al. (2006a) argue that stereotypes activate indexed social exemplars, and the activation spreads to the distributional properties of the phonetic exemplars.

Exemplar models have been widely adopted in theories of speech perception (e.g., Johnson, 1997; Goldinger, 1998; Pierrehumbert, 2001), and provide an elegant theoretical ground for the association between linguistic variants and speaker information. In Exemplar Theory, each individual's experiences with various phonetic realizations of a lexical item are registered in episodic memory as phonetic exemplars and encoded together within the representation.[4] Perceiving an utterance involves mapping the incoming signal to an existing exemplar based on the similarity between them. Early findings show that words are recognized more accurately and quickly when spoken by the same individual that listeners previously heard producing those words, rather than when produced in a novel voice (Mullennix et al., 1989; Palmeri et al., 1993; Goldinger, 1997). The priming effect of talker specificity was also observed when tested with newly learned nonsense words (Creel & Tumlin, 2011). These studies indicate that information about the voice characteristics of a particular speaker is stored in memory, and the perceptual link between a particular speaker and a word leads to a processing advantage.

As shown by the body of sociophonetic literature introduced above, the talker specificity effect is not limited to listeners' sensitivity to individual speakers in experimental sessions, but listeners are also sensitive to the generalized distribution of phonetic variation over particular social categories. This sort of generalization processes can be understood in the context of a hybrid account of exemplar models (e.g., Pierrehumbert, 2003; Pitt, 2009; Pinnow & Connine, 2014), which views that supposing either a single invariant abstract representation (e.g., Gaskell & Marslen-Wilson, 1998) or multiple representations stored in episodic memory (e.g., Johnson, 1997; Goldinger, 1998) is not sufficient to explain how novel variants are recognized as words. The hybrid models do posit multiple representations for a given word, but the representations are abstract, i.e., underspecified for certain features, and weighted by frequency of each variant, allowing room for generalization over variant properties (Pinnow & Connine, 2014). Pierrehumbert (2003) argues that individuals improve recognition of variant forms by developing multiple levels of representations during lifetime, and the representations are generalized and refined based on the statistical regularities found through additional experience. Exemplars stored in long-term memory then can either establish a representation that is closely associated with an existing category, or be generalized into an existing category, if the incoming speech signal is sufficiently similar in its phonetic characteristics (Pierrehumbert, 2001; Bybee & Cacoullos, 2008). Through this generalization process, frequently activated exemplars are

---

[4] These representations can be at any level of linguistic analyses, including segments, words, and even morpho-syntactic level (Bybee & Cacoullos, 2008). However, lexical representations are the main interest of this paper.

associated with a variety of episodic memories, forming dense exemplar clouds, which in result lead to processing advantages (Pierrehumbert, 2001).

In an exemplar-based account, phonetic exemplars are indexed to social information about the speaker, and the indexical information is acquired in an early age (Foulkes & Docherty, 2006). Once phonetically-rich exemplars are activated during perception, the activation automatically spreads to social information indexed to the exemplar (Drager & Kirtley, to appear). Reversely, the activation of social information spreads to phonetically-rich exemplars of words (Hay et al., 2006a; Hay et al., 2006b). Therefore, activation of social information expedites lexical access when the social and lexical information are linked, and it even biases perception when the perceived sound acoustically matches well with the existing mental representation in the listener's mind.

There are also non-exemplar-based models that emphasize that social exemplar activation does not necessarily rely heavily on token frequency. Sumner and colleagues propose a model in which acoustic patterns are mapped to both linguistic representations and social representations in tandem. Sumner and Samuel (2009) show that listeners' experience strongly affects recognition and representation of r-ful or r-less variant forms of American English. In contrast, Sumner and Samuel (2005) demonstrate that a socially-idealized phonetic variable, e.g., [fluːt] for *flute*, receives more attention and is encoded more strongly than a frequent but non-idealized form, e.g., [flu̠ʔt˺], and thus the clusters of exemplars for that variable can be as robust as those of the exemplars for the frequent counterpart. This is what Sumner et al. (2014) call "socially-weighted encoding." In Sumner et al.'s model, a frequent form benefits from typicality during the processing of lexical representation, but the socially-weighted encoding of the frequent form is not as strong as an infrequent socially-idealized form. Therefore, both of the forms can trigger similar amounts of priming effect and be recognized faster than a form that is neither frequent nor socially idealized. In addition, even an infrequent and non-idealized form e.g., [flu̠ʔ], can benefit from typicality if it has a similar acoustic value. Sumner et al. (2014), therefore, propose a dual-route mechanism, in which the encoding of speech to lexical processing interacts with the encoding of speech to social features.

Models using Bayes's theorem (e.g., Norris, 2006; Norris & Mcqueen, 2008) also predict that comprehension is affected by speaker information, although what they predict about the timing of the social effect may differ from exemplar-based models and other connectionist models (we return to the timing issue in Section 5). Not assuming activation triggered by similarity, they posit that lexical access approaches the optimal representational node by combining the perceptual evidence with knowledge of the prior probabilities of words (Norris & Mcqueen, 2008). Visually-provided information about the perspective and characteristics of speakers is integrated to predict sentence-level meaning (Kamide et al., 2003; Hanna & Tanenhaus, 2004). When listeners are provided with voice-based speaker information in just a few words preceding a target word, linguistic meaning of the target word is evaluated against the pragmatic knowledge, and the integration occurs at the same time and in a shared brain region with the decoding of semantics (Van Berkum et al., 2008; Tesink et al., 2009). Taken together, in the Bayesian account, speech recognition makes use of listeners' expectancy for what the speaker is going to say based on the contextual information.

To summarize, there is ample evidence from empirical and theoretical background that speech recognition is affected by associations between linguistic variants and social categories. Crucially, Walker and Hay (2011) demonstrated that congruence between speaker age and word age facilitates word recognition. They argue that exposure to the words in use is the source of the observed effect, but they do not rule out that stereotypes can also

influence lexical access. This raises the question: Do stereotypes about words have an effect on lexical access and, if so, do influences from stereotypes add to or wash out effects from exposure? We explore this question through a modified version of their experiment, with word ages determined in two folds. First, usage age is determined based on a cross-age comparison of self-reported measures of how frequently Korean speakers verbally produce each word. Second, stereotypical word age is measured based on native speakers' judgment on whether younger or older people are more likely to use each of the words. The two types of ratings are used as predictors of recognition accuracy and response times, and their effects are compared. We hypothesized that associations between linguistic representations and speaker information are not only referenced to distributional properties but also reinforced by word-stereotypes.

Additionally, the vast majority of work demonstrating an effect of social information on perception is based on English. Replicating their results in Korean, a language that is unrelated to English and interacts with its own sociolinguistic constructs, would help demonstrate that the effect is widespread and therefore needs to be accounted for in our linguistic models.

The Korean lexicon is especially suitable for testing the two types of associations because jargons of the younger speakers are massively used in spoken language as well as in text communication. One phonological reason for that relates to how acronyms, which takes a large portion of the jargons, are created in Korean. Virtually all initialized words in Korean are acronyms rather than initialisms, in that the outcomes are phonotactically legal, and thus pronounceable as words. Unlike initialisms in English, they are created by combining initial syllables, not initial segments, of each word. For example, the noun phrase, *ppesu khatu chwungcen* 'charging a bus card' is contracted to *ppe.kha.chwung*. The output is not just pronounceable but also contains phonological information at a syllable level which helps learners predict what words are shortened, and thus contributes to proliferation of young words in spoken language.

Another reason comes from a social aspect. South Korea has accomplished rapid economic growth through the postwar era. Since the time when baby boomers (who are now over 50) were young, the national industry has gradually shifted from rural-based agriculture to an urban-focused capitalist system. New values of the society arose in step with the socio-structural changes. Community ethics back then cherished the values of extended families and community-centered goals, whereas individualism is widespread among the younger people of the current society. What followed is diversity in language, strengthening solidarities and identities of social groups, including those of age groups. Moreover, as one of the countries with the fastest internet connection speed (Stenovec, 2015), South Korea is also ranked on top in smart phone ownership and internet usage rate (Poushter, 2016). With such technological infrastructure, the spread of young words is accelerated as a form of internet slang. Some of these impacts on the language are reflected in selection of the lexical items (see Section 3.1).

3 Method
3.1 Lexical Stimuli

The stimuli were made up of 384 words and 384 non-words. Word stimuli were chosen based on an online survey, through which both types of word ages were measured. Survey respondents were 80 Seoul Dialect speakers living in the Seoul Metropolitan area, recruited from a separate population from the lexical decision experiment participants. There were 42 older respondents (aged from 50 to 71) and 38 younger respondents (college students aged from 18 and 25).

In the survey, 340 words (2–4 syllables) were presented in written form with three questions per word. For items with non-standard phonetic forms, surface forms were

presented rather than their standard orthography (e.g., 빠다 (*ppata*[5]), instead of 버터 (*pethe*) 'butter'). The survey items were chosen by the author, including 116 potentially young words, 179 potentially old words, and 45 age-neutral words. First, potentially young words were selected from three lexical categories: (1) words that are semantically associated with campus life (e.g., *kkwacampa* 'a jacket worn by students in a department to mark their sense of solidarity'[6]), (2) coined words that originated mainly from computer-mediated communication (e.g., *notap* 'no solution', a word made of *no* in English and *tap* 'a solution' in Korean, sarcastically describing someone's poor performance or stupid behavior), and (3) words for new concepts (e.g., *pheyisupwuk* 'Facebook'). Second, potentially old words consisted of four categories: (1) words that are semantically associated with rural life (e.g., *cangtoktay* 'a platform [in a traditional house] used to store crocks of sauces and condiments'), (2) vanishing terms (e.g., *kwukminhakkyo*, an antiquated term for 'an elementary school'), (3) kinship addressing terms that had been frequently used under an extended-family system (e.g., *ansaton* 'one's daughter-in-law's [or son-in-law's] mother'), and (4) loanwords that had been imported from Chinese and Japanese before the independence from the colonization by Japan in 1945, and then became discouraged by the so-called "National Language Purification Policy" (e.g., *syassu* 'a shirt,' *pwullanse* [佛蘭西] 'France'). Lastly, potentially neutral words were ones that are not highly frequent, not apparently associated with either age group, and do not belong to any of the potentially-aged categories (e.g., *kkatalk* 'reason,' *yutaykam* 'a sense of fellowship,' *sonkalakcil* 'finger-pointing').

  The 340 words in the survey were divided into two sets of 170 words. Forty respondents (7 older males [OM], 14 older females [OF], 11 younger males [YM], 8 younger females [YF]) answered questions for one set (58 young, 90 old, 22 neutral words), and the other 40 respondents (8 OM, 13 OF, 7 YM, 12 YF) answered for the other set (58 young, 89 old, 23 neutral words). The first question in the survey asked how often the respondents see or hear other people using each word to measure word familiarities. Respondents were asked to choose their answer from four statements about the degree of exposure, and their answers were coded as *reported-exposure scores*, from 0 ("I have never seen or heard the word, and I don't even know the word.") to 3 ("I see or hear it pretty often, at least once a week or more frequently."). The second question asked how often they use the word in real conversation to measure frequencies of verbal use. Respondents' answers were coded as *reported-usage scores*, from 0 ("I have never spoken the word.") to 3 ("I use the word pretty often, at least once a week or more frequently."). The third question was asked only of those who did not choose 0 in the first question (excluding those who do not know the word): "Between younger people (10s–20s) and older people (60s and above), which do you think use the word more frequently?" Respondents chose from five statements, and *stereotype scores* were coded from –2 ("The young use it much more frequently.") through 0 ("Both use similarly frequently."), to +2 ("The old use it much more frequently.").

  Two predictors of word age were drawn for each of the 340 words, based on the ratings for the last two questions. First, the *young respondents' stereotype score* (ST score) was obtained by calculating the mean of the stereotype scores for each word that only the young survey respondents rated. Only responses from the young respondents were used

---

[5] Phonemic transcriptions in this paper are based on "The Yale Romanization System."

[6] Definitions of the young word examples are based on an online dictionary of Korean neologism, Naver Open Dictionary (http://kin.naver.com/opendic/index.nhn), except *kkwacampa*, which was not found in any of the available dictionaries. Definitions of old word examples are based on the ET-house Neungyule Korean-English Dictionary (http://www.et-house.com/pages/dictionary/ko_dic.asp) and the Standard Korean Dictionary of National Institute of Korean Language (http://stdweb2.korean.go.kr/main.jsp).

because they are the same age as participants in the lexical decision experiment (reported in Section 3.5). Second, the *usage-age score* (UA score) was calculated for each word by subtracting the mean of the younger respondents' reported-usage scores (in the second question) from the mean of the older respondents' reported-usage scores. Therefore, a high (positive) UA score indicates that the word is used more by older people, and a word with a low (negative) score is used more by young people.

It is worthwhile to note that the UA score is based on self-report and that the respondents' answers may have been affected by the stereotypes that they were also asked about. This contrasts with Walker and Hay's test of token frequency, which was corpus-based. This could not be done for the current study because many of the items do not appear in the Sejong spoken corpus (or any other spoken/written corpora), nor do they appear in standard Korean dictionaries. However, the introspective subjective measures of individuals' lexical knowledge based on experience, such as ratings of subjective familiarity and frequency, have been attested to be an efficient method to capture the characteristics of word representations based on individuals' experience (Kuperman & Van Dyke, 2013). Kuperman and Van Dyke argue that subjective measures, as opposed to objective measures based on corpora, can properly represent the word representations that are shaped by individuals' experience. Therefore, despite being a subjective parameter, the UA score likely portrays an accurate relative representation of word frequency based on individuals' representational differences.
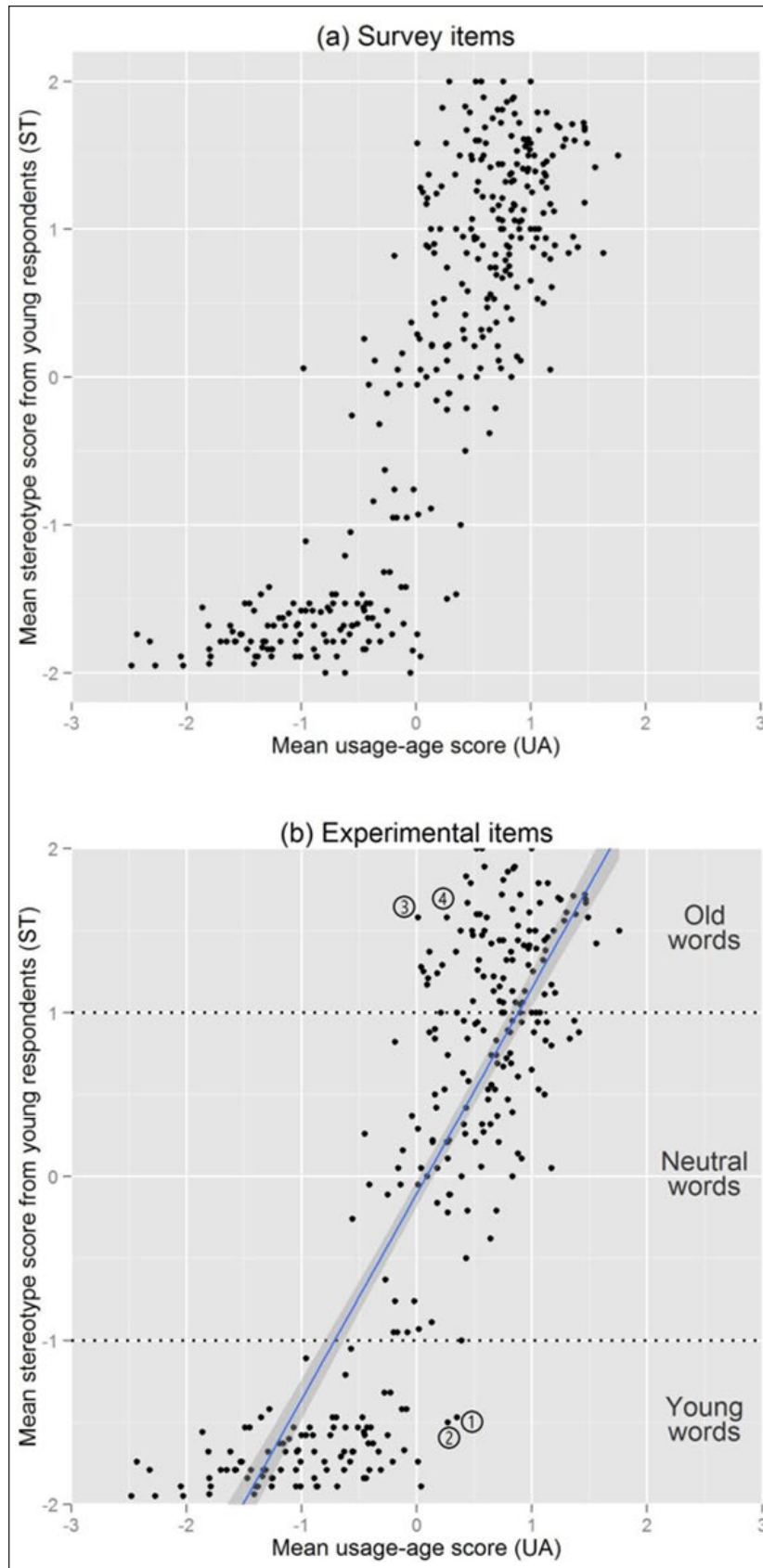
Figure 1(a) shows the distribution of ST and UA scores for all of the 340 survey items, from which 52 items were excluded based on two criteria. First, nine words were excluded due to low mean reported-exposure scores for the young survey respondent group (under 0.47/3.00), to prevent high error rates in the lexical decision experiment. Second, 41 items were removed from densely populated regions in **Figure 1(a)**, so that items were not centered around a particular range of ST and UA scores. **Figure 1(b)** shows the distribution of the selected experimental items, where the boundaries between each word age group are marked by dotted lines. Old words were selected from the words at or above the ST score of +1. Young words were selected from the words at or below –1. Neutral words were selected from the words in between (N = 96 for each category).

Unsurprisingly, the ST scores and UA scores are highly correlated in a Kendall's tau correlation test (τ = .71, p < .001). However, there are some noticeable trends in the distribution of data points.[7] The distribution of young words is closer to the ceilings of the scale, i.e., –2 in ST and –3 in UA, than old words both in ST and UA scores, and neutral words are centered around a region with relatively high ST and UA score, i.e., not around (0, 0). There are some words that have low ST scores (i.e., they are associated with young people), but relatively high UA scores (i.e., more older people reported using them). These include *khaphwuchino* 'cappuccino' and *tikha* 'a digital camera,' marked by ① and ②, respectively in **Figure 1(b)**. Conversely, some words have high ST scores but low UA scores (e.g., *ipoke* 'hey' and *tapang* 'a tea house,' marked by ③ and ④, respectively). Examples of words with correlated ST and UA score are a young word, *notap* 'no solution,' and an old word *kwucwa*, a word imported from Japanese, meaning 'a bank account.'

In addition to these 288 words, 96 filler words were chosen from the most frequent words in the Sejong spoken corpus. **Table 1** summarizes the word and non-word stimuli used in the lexical decision experiment. A list of young and old words can be found in the Appendix.

---

[7] Note that this data is based on means of ordinal data and are not normally distributed.

Figure 1: Distribution of stereotypical word age (ST) and usage age (UA) scores for **(a)** the survey items and **(b)** the experimental items. In **(b)**, circled numbers indicate examples of items with non-matching UA and ST scores. The two dotted lines are boundaries of word age categories, and the solid line represents a linear regression in which ST is regressed by UA.

| Number of syllables			Two	Three	Four	Total
Words	Critical	Young	48	36	12	96
		Old	48	36	12	96
		Neutral	35	41	20	96
	Filler		85	10	1	96
Non-words			216	123	45	384

Table 1: Summary of stimuli.

Non-word stimuli were intended to be as word-like as possible to yield sufficient processing time to observe the effects. They were created by changing one or two segments of existing Korean words that were not used as word stimuli. Base words were chosen from the same word categories used for potentially old, young, and neutral words. All non-word stimuli are phonotactically legal but do not appear in the Standard Korean Dictionary of National Institute of Korean Language (http://stdweb2.korean.go.kr/main.jsp).

3.2 Auditory Stimuli

Four speakers were selected through a speaker norming test. The purpose of the norming test was to normalize voice age across speakers and minimize effects of sociolinguistic information contained in the voice other than perceived voice age. Sample recordings of 26 speakers (7 OM, 7 OF, 6 YM, and 6 YF recruited from the Korean community in Hawai'i) were rated by seven raters who were naive to the research purpose (2 males and 5 females, aged from 31 to 37). Raters were instructed to rate social characteristics they attributed to each speaker based on recordings of six word stimuli (two words in each word age category). Raters were asked about age, dialect, socio-economic status, education level, and how likely the speaker is to use young language. Based on the results of the rating task, 4 speakers were selected (OM: 60 years old/rated as 53.14, OF: 75/63.86, YM: 26/33.14, YF: 22/20.71). Of the original 26 speakers, the final 4 were the ones in their respective age group whose ratings most closely met five criteria: someone who is (1) a fluent Seoul Dialect speaker, (2) in middle or upper-middle socio-economic status, (3) between 18 and 25 years old for young speakers and over 50 for old speakers, (4) currently a college student for young speakers and at least a high school graduate for old speakers, and (5) likely to use young language for younger speakers and unlikely to do so for older speakers.

Stimuli were recorded by each speaker in a sound-attenuated booth at the University of Hawai'i. A portable Tascam DR–7 recorder was used to record with a mono, 32-bit, 44,100 Hz sampling rate setting. Each word was produced twice, preceded by a carrier phrase, *ipen tanenun* 'This word is …,' and the critical words were excised from the carrier phrase using Praat (version 5.4.04, retrieved 29 January 2015 from http://www.praat.org/). Due to the lack of lexical stress in Seoul Korean, it was necessary to control for prosodic differences across speakers and within speakers. All bisyllabic items were spoken in a high tone on the first syllable (H+), followed by a low IP-final boundary tone (L%). Trisyllabic and quadrisyllabic items were spoken in either of two types of tonal frames, following Jun's (2011) Revised Intonation Model of Seoul Korean. For words that begin with an aspirate or tense consonant, or /h, s/, the first two syllables were produced in high tones (H+H), and a low IP-final boundary tone (L%) was assigned starting from the third syllable. Otherwise, the first two syllables were produced in a low and a high tone (L+H), followed by L% (see **Figure 2**). In the recording session, speakers listened to pre-recorded samples in the author's voice and were directed to produce each word similarly, imitating the rhythm and pitch of the samples as closely as possible. Speakers were corrected immediately by the author when mispronunciation occurred.
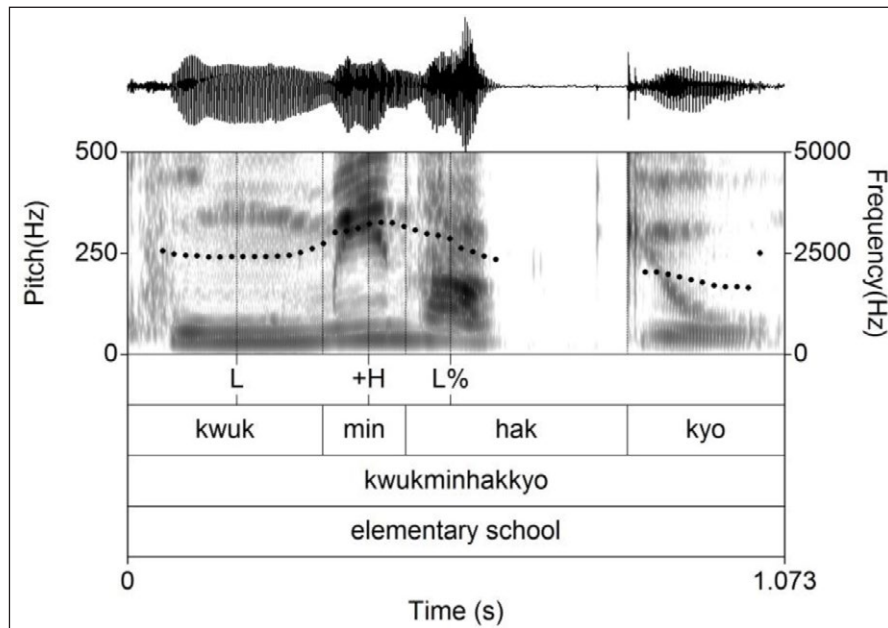
Figure 2: Recording example of a quadrisyllabic word, kwukminhakkyo 'an elementary school,' spoken by the YF speaker.

3.3 Procedure

A lexical decision experiment was conducted in a sound-attenuated booth at the University of Hawai'i. All stimuli were spoken in isolation in order to prevent any sentential or pragmatic influence on the decision of the lexical status. Participants heard auditory stimuli and were asked to press a button as quickly and accurately as possible with their dominant hand if they thought they heard a Korean word that they understood, and press another button with the other hand if not. The two buttons were placed at the far ends of a Cedrus response box. Participants were asked not to move their hands off the response box as much as possible during the experiment. Participants were also informed that a 'word' in this experiment referred to any variety of word forms that Korean speakers use, including loanwords, non-standard phonetic variations in dialects, and coined words. In practice trials, feedback was provided on the computer monitor screen to inform participants whether their response was correct. Throughout the experiment session, the screen only showed "Experiment in progress" on a white background. When participants took longer than 1,300 ms to respond, a warning message appeared on the monitor, asking them to respond quickly. It took about 45 minutes on average to complete the whole experiment session, which was blocked by four speakers. Participants were able to choose to take a break between blocks, and there was a mandatory break after the second block, during which a two-minute video presented natural scenery from Hawai'i. The experiment session was followed by an exit survey, where participants heard two quadrisyllabic word stimuli in each of the four speakers' voices and rated perceived social information about each speaker (e.g., voice age, socio-economic status, and education level).

3.4 Design

The experiment session was divided into four blocks, with one speaker per block. Stimuli were counterbalanced by speaker age while gender was controlled. That is, if one group of participants heard an item in the OM speaker's voice, then the other group heard that item in the YM speaker's voice, and vice versa for the OF and YF speakers. The order of speakers was also counter-balanced by speaker age but not by gender. One group heard

the OM and OF speakers first, followed by the YM and YF speakers. The other group heard the younger speakers first, followed by the older speakers. Participants were assigned randomly to one of these four (2*2) lists of stimuli.

3.5 Participants

For the lexical decision experiment, 35 native speakers of Korean (25 females and 10 males) were recruited from a pool of exchange students at the University of Hawai'i. All but two of them (one 28-year-old UH student and one 29-year-old kindergarten teacher from Korea) were 20–25-year-old registered students at colleges in the Seoul Metropolitan area at the time of participation. All participants listed Seoul Dialect as their first or second most frequently used dialect. Participants were paid for participation.

4 Results

4.1 Accuracy

Among the 26,880 tokens in total, there were 23,040 correct responses. The accuracy rate was 86.67% for words, and 84.75% for non-words. Accuracy rates for words in each level of the word age condition and speaker age condition are provided in **Table 2**.

For each category of words, words spoken by younger speakers were recognized more accurately than older speakers (87.98% for younger, 85.37% for older, averaging all categories). Regardless of speaker age, filler words, i.e., frequent words in the Sejong corpus (93.01%) and young words (91.49%) were recognized more accurately than age-neutral words (85.71%) and old words (76.49%). The number of items for which less than 50% of the participants recognized correctly was 0 (filler), 1 (young), 4 (age-neutral), and 14 (old), among 96 items in each category. Importantly, accuracy rate was slightly higher when word age and speaker age matched (84.64%) than when they mismatched (83.33%). Although old words were recognized better for younger speakers (77.14%) than older speakers (75.83%), young words were also recognized better when spoken by younger speakers (93.45%) than older speakers (89.52%).

To test whether there is a significant effect of stereotypical word age and speaker age on the number of correct responses, a binomial mixed effects model was fit to the binary accuracy data of the critical items (excluding fillers and non-words) in R (version 3.2.2, retrieved 13 October 2015 from https://cran.r-project.org/). An interaction between binary speaker age (older or younger, deviation coded) and the ST scores (treated as continuous) was included as a fixed effect in the model. By-item intercepts, by-participant intercepts, and by-participant slopes of speaker age, ST score, and the interaction between them were included as random effects. As shown by the positive estimated coefficient for speaker age in **Table 3**, words spoken by younger speakers were significantly more frequently recognized correctly than words spoken by older speakers ($p < .001$). There was also a significant main effect of ST score; as ST score increased, participants were less likely to respond correctly ($p < .001$). Importantly, there was a significant interaction between speaker age and word age. When words were heard in younger speakers' voice, in comparison with older speakers' voice, participants were less likely to respond correctly for words with higher ST scores ($p = .002$).

Speaker Age	Filler	Word Age			Total
		Young	**Neutral**	**Old**	
Younger	93.75%	**93.45%**	87.56%	**77.14%**	87.98%
Older	92.26%	**89.52%**	83.87%	**75.83%**	85.37%
Total	93.01%	91.49%	85.71%	76.49%	86.67%

Table 2: Accuracy rates for words by word age and speaker age.

For comparison, the effects of usage-based word age and speaker age on accuracy were examined in a separate binomial mixed effects model, by replacing the predictor of word age, i.e., ST score, with UA score. All other predictors tested in this model were identical with the model predicted by ST score. As shown in **Table 4**, there were significant main effects of speaker age and UA score in the same directions as in the previous model. That is, words spoken by younger speakers were more likely recognized correctly ($p < .001$), and words with high UA scores were less likely recognized correctly ($p < .001$). The interaction between speaker age and word age was also significant, indicating that recognition accuracy was decreased when words with high UA scores were spoken by younger speakers ($p = .028$).

4.2 Reaction Times

Reaction times (RT) were measured only for the real word trials with correct responses (11,649 tokens) from the onset of a target word to the response (button press). Data points below 300 ms and over 5,000 ms were removed first (N = 7), and then responses over three standard deviations from the mean by participant were removed (N = 133). This left 11,509 tokens for the analysis, excluding 140 outliers (1.20% of the total). Raw RTs were transformed into residual RTs to observe the effects of the experimental conditions, with the effect of word duration on individual listeners being controlled (illustrated in **Figure 3**). To calculate residual RTs, a linear mixed effects model was fit to the raw RTs, with word duration as a fixed effect and participant as a random effect. Residual RTs were obtained by subtracting RTs predicted by the model from the raw RTs.[8]

Means of residual RTs in **Figure (3a)** indicate that there are main effects of word age and speaker age; word recognition occurred faster when the participants heard young words and younger speakers. Also, the difference of residual RTs between younger speakers and

	Estimate	Std. Error	t value	p value
(Intercept)	2.518	0.144	17.469	<.001
Speaker age=young	0.244	0.066	3.720	<.001
ST score	–0.513	0.078	–6.578	<.001
Speaker age=young : ST score	–0.115	0.037	–3.070	.002

Table 3: Summary of results of binomial mixed effects model fit to accuracy with the interaction of speaker age and ST score as a fixed effect. A high ST score indicates that the word is highly associated with older speakers.

	Estimate	Std. Error	t value	p value
(Intercept)	2.559	0.148	17.339	<.001
Speaker age=young	0.227	0.065	3.515	<.001
UA score	–0.633	0.119	–5.296	<.001
Speaker age=young : UA score	–0.115	0.052	–2.198	.028

Table 4: Summary of results of binomial mixed effects model fit to accuracy with the interaction of speaker age and UA score as a fixed effect. A high UA score indicates that the word is highly associated with older speakers.

---

[8] Residual reaction times are often used in self-paced reading analysis (e.g., Ferreira & Clifton, 1986), to adjust raw reading times for differences of individual participants' sensitivity to the length of the words in the critical region (e.g., number of syllables, letters, or moras). In the model that predicts the effect of word length, only participant, but not item, is included as a random effect because the effect of word length is intrinsically dependent on an individual's sensitivity, but not on the characteristics of words.
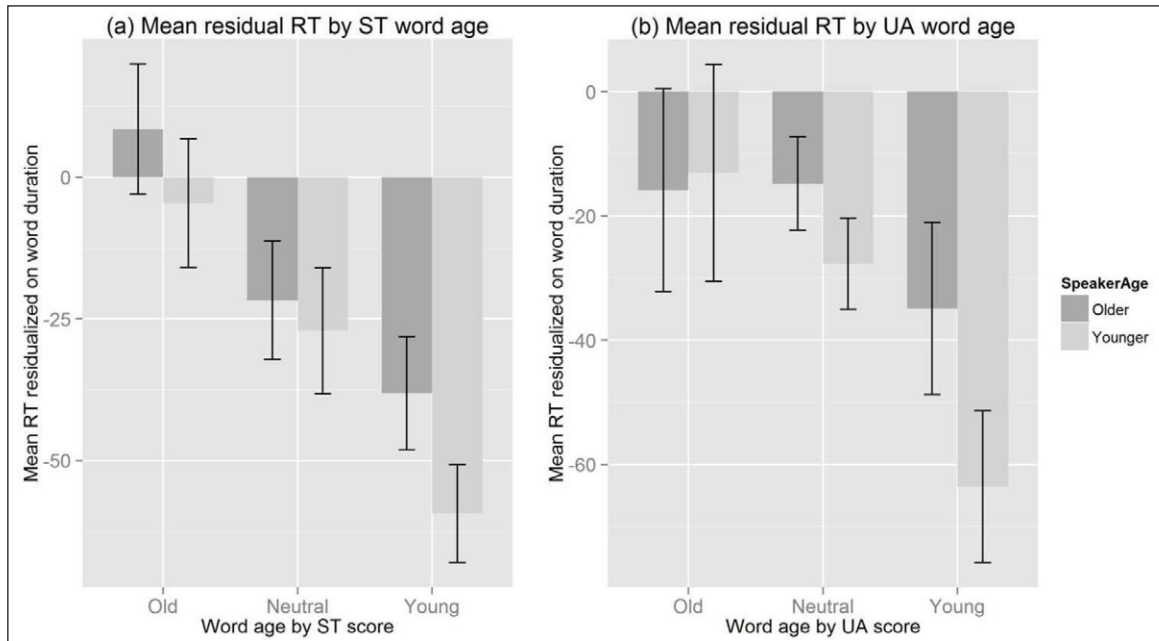
Figure 3: Means of residual RTs conditioned by word age and speaker age. Word age is categorized into old, neutral, and young words, **(a)** based on the ST score, and **(b)** based on the UA score. Raw RTs were transformed to residual RTs on the y-axis, by subtracting the by-participant effect of word duration from raw RTs. Error bars represent 95% confidence interval of means by participants.

older speakers was greater when the word age was young. To test whether these effects are statistically significant, a linear mixed effects model was fit to raw RTs (not residual RTs) of the critical items in R. Fixed effects in the model were word duration (as a control variable) and an interaction between binary speaker age (older or younger, deviation coded) and ST scores (treated as continuous). By-item intercepts, by-participant intercepts, and by-participant slopes of speaker age, ST score, and the interaction between them were included as random effects. Test block order and trial order were also tested but only those factors that reached significance were included in the final model in addition to the test variables.

As shown by the negative estimated coefficient for speaker age in **Table 5**, RTs were decreased when the words were spoken by younger speakers than by older speakers, although the effect was not significant ($p = .112$). There was a significant main effect of ST score; as ST score increased, participants recognized the words more slowly ($p < .001$). Also, there was a significant main effect of word duration; as the duration of a word increased, RTs increased significantly ($p < .001$). Importantly, there was a significant

	Estimate	Std. Error	*t* value	*p* value
(Intercept)	713.774	23.145	30.839	<.001
Speaker age=young	−7.607	4.783	−1.591	.112
ST score	20.010	4.597	4.353	<.001
Duration	0.612	0.024	25.671	<.001
Speaker age=young : ST score	3.068	1.542	1.990	.047

Table 5: Summary of results of mixed effects model fit to reaction times with ST score as a fixed effect. A high ST score indicates that the word is highly associated with older speakers.

interaction between speaker age and word age (continuous ST score). When words were heard in a younger speaker's voice, in comparison with an older speaker's voice, participants were slower to recognize words with higher ST scores ($p = .047$).

Reaction times from the endpoint of a word to the response were also analyzed, and similar patterns were observed, but the interaction between ST score and speaker age was marginal ($p = .070$). Similar patterns were also observed when continuous speaker age, i.e. the perceived ages of each speaker rated by participants in the exit survey (see Section 3.3), was included as a fixed effect in the model, instead of binary speaker age. However, the interaction between ST score and speaker age was marginal ($p = .053$). Also, the interaction was not significant either in accuracy ($p = .283$) or in RT ($p = .139$), when by-item random slopes were added to the ST score models. This seems to be due to specific items rather than the group-level factor. This model indicated high by-item variability on the random intercept (variance $= 7072.94$, $SD = 84.101$), on the coefficients for ST score (variance $= 779.54$, $SD = 27.920$), and on speaker age (variance $= 1239.41$, $SD = 35.205$).

For comparison, the continuous UA score was also tested as a fixed effect (instead of ST score) in a separate linear mixed effects model fit to raw RTs of the critical items, by replacing the predictor of word age, i.e., ST score, with UA score (see **Table 6**). All other predictors tested in this model were identical with the model predicted by ST score. As shown in **Table 6**, participants recognized words faster when the words were heard in younger speakers' voices than in older speakers' voices. However, the effect was marginal ($p = .096$). There was a significant main effect of UA score; as UA score increased, participants recognized the words more slowly ($p < .001$). Also, there was a significant main effect of word duration; as the duration of a word increased, RTs increased significantly ($p < .001$). Importantly, the interaction between speaker and word age was only marginal ($p = .053$) when continuous UA score was included in the model as a predictor of word age.

4.3 Model Comparison

The above analyses demonstrate the effect of either stereotype-based or usage-based word age. Model comparisons were performed to test whether one of the two effects on RTs can be isolated from the other. The first comparison tested whether the residual sum of squares are significantly reduced, when ST score residualized on UA are added as a predictor variable in addition to UA score[9]. First, a preliminary linear regression model was fit to predict the ST score with UA score. The residuals from this model, or residualized ST, constitutes a variable of ST scores independent from the contribution of UA scores to RTs. Then, residualized ST was added as a predictor into the aforementioned model

	Estimate	Std. Error	*t* value	*p* value
(Intercept)	717.093	23.137	30.993	<.001
Speaker age=young	–7.982	4.793	–1.665	.096
UA score	25.665	6.691	3.835	<.001
Duration	0.605	0.024	25.267	<.001
Speaker age=young : UA score	4.156	2.151	1.932	.053

Table 6: Summary of mixed effects model fit to reaction times, with UA score as a fixed effect. A high UA score indicates that the word was more often reported as being used by older speakers than younger speakers.

---

[9] I appreciate an anonymous reviewer's suggestion for this. This method is used in psycholinguistic analyses to evaluate the effects of correlated predictors (e.g., Kuperman et al, 2010; Cohen-Goldberg, 2012).

with UA score (**Table 6**). This new model was compared to the original model without residualized ST, using analysis of variance (ANOVA). The results revealed that the residualized ST improved the fit of the model with UA score only ($\chi 2(1) = 4.76$, $p < .029$), and a significant main effect of residulalized ST remained in the new model ($\beta = 19.005$, $SE = 8.722$, $p = .029$). The second model comparison was performed for UA score residualized on ST to dissociate the effect of UA from ST. The results revealed that the residualized UA neither improved the fit of the RT model that included ST score (**Table 5**) ($\chi 2(1) = .03$, $p = .856$), nor remained significant ($\beta = 2.244$, $SE = 12.739$, $p = .860$). The two comparisons, therefore, indicate that ST score has an effect above and beyond an effect of UA score, but not vice versa.

5 Discussion

As a summary of the results, both accuracy rates and recognition reaction times were improved when speaker age matched either type of word ages. When word ages were predicted by ST score, the interaction between voice age and word age significantly affected both accuracy rates ($p = .002$) and reaction times ($p = .047$). In contrast, when a frequency-based rating was used as a predictor of word age, the effect was significant in accuracy rates ($p = .028$) but marginal in reaction times ($p = .053$). Finally, residualized ST scores remained a significant main effect when added as a predictor in addition to UA scores, improving the fit of the model only with UA scores, while residualized UA scores did not.

These results corroborate that phonetic cues for talkers' social categories are utilized in speech perception (e.g., Johnson et al., 1999; Niedzielski, 1999; Hay et al., 2006b; Staum-Casasanto, 2008; Drager, 2011). The results also agree with those of Walker and Hay (2011) that lexical access is facilitated by matched social information between words and speakers, providing evidence from Korean, in which the social and phonological constructs are different from those of English. The hybrid approach to exemplar modeling (Pierrehumbert, 2003; Pitt, 2009; Pinnow & Connine, 2014) provides a basis for how it is possible that listeners are sensitive to a generalized set of exemplars that associate a certain social group. Under the assumptions of the exemplar-based approach with social indexing (Foulkes & Docherty, 2006; Hay et al., 2006b; Drager & Kirtley, to appear), age cues in the signal automatically activate relevant episodic memories and age-matching social indices. The activation spreads to the distributional association of words via social indices, guiding to the indexed lexical representation. As a result, recognition is facilitated when the social exemplars activated by the voice match well with the exemplars stored in the target representation. It should be also noted that the core predictions of exemplar models are concerned with representations indexed to individual's subjective experience. In this sense, although counts of word occurrences may be an efficient measure for individuals' experience, the finding that recognition proficiency is predicted by self-reported production frequency does not stray from the predictions of the exemplar models.

On top of that, the results suggest that social exemplar activation is also closely related to the stereotypical association. An additional effect of stereotypical associations was found over that of experience. This finding can be compared with Walker and Hay's (2011) findings, in which the effect of relative word frequency was present, whereas linguistic stereotypes for lexical items did not show a significant effect. Although both experiments show the processing advantage driven by activation of a social link between lexical representations and speaker information, the two are contrastive in terms of on what level of processing the cognitive association is realized.

The apparently incompatible contrast can be interpreted as a consequence of different methods used in the two experiments. The experiment design in this paper made age stereotypes more salient than they were in Walker and Hay (2011). Word-stereotypes

were initially delineated by native speakers' conscious metalinguistic judgments, and then items were selected based on them. However, Walker and Hay's items were selected based on relative frequency data. Consequently, although the frequency-based word ages had a significant effect on the unconscious and automatic activation of social exemplars, the age distinction was less likely to emerge above a conscious level of association in their post hoc survey.

In addition, the UA and ST scores plotted in **Figure 1** exhibit a flooring effect in general. That is, many young words were reported to be almost exclusively used by, and highly stereotypically associated with, younger people. As a result, the listeners, who were all in their 20s, were able to recognize the young words correctly as often as the filler items that are highly frequent words in the Sejong corpus (see **Table 2**). The flooring effect seems to be due to the inclusion of a large number of non-standard items, e.g., internet slang. Only 14 of the 96 young words are currently registered in the Standard Korean Dictionary of National Institute of Korean Language[10]. In comparison, 87 items among the 96 old words appear in the dictionary, but many of them are accompanied in the dictionary with an annotation that those forms are inappropriate and need to be replaced with standard forms. Therefore, the old words were also salient, so the distinction between age groups was salient enough to attract participants' awareness, maximizing the effect of age stereotypical associations.

In sum, the findings from this paper and Walker and Hay (2011) are not incompatible, but shed light on different aspects of perceptual associations. What Walker and Hay discovered is that words' distribution affects lexical recognition automatically, without listeners' explicit awareness of the association. In contrast, this study finds that stereotype is an additional predictor of lexical access, of which effect appears to be stronger than an effect of distribution, at least when listeners are exposed to words on a wide spectrum of age-related variability.

Our view on stereotypes is that they are formed not just by frequent use by a particular group of speakers, but also filtered through another layer of social and linguistic indices, leading to a conscious level of awareness for the variable. These indices may involve contextual domains (e.g., internet slang or traditional kinship terms), social changes (e.g., new concepts or vanishing terms) and semantic priming (e.g., rural life or campus life). Put on the additional layer, stereotypes can override the known effect of frequency. Therefore, lexical representations are formed and shaped by experience and distributional properties, but the social indices of the representations are reinforced by stereotypes.

Stereotypical associations should be considered in connection with the capacity and limitations of the frequency-based mechanism. Apparently, stereotypes may be directly indexed to representations as a different route than distributional properties. However, we do not abandon the possibility that the additional effect of stereotype is limited to certain words. In addition, what stereotypes actually do may be nothing more than magnifying the activation of distributional indices of word representation (as suggested by Hay et al., 2006a; 2006b).

In this vein, models of speech perception should take into account the interaction between the distributional factor and the ideological distinctions of social groups, which is in no way negligible amid the formation of linguistic stereotypes. One such issue relates to the roles of and relation between salience and encoding strength in perceiving sociolinguistic variables. The salience of the word ages in this study may have drawn listeners'

---

[10] Most of the young words can only be found in online dictionaries of neologism, such as the Naver Open Dictionary (http://kin.naver.com/opendic/index.nhn). In contrast, Walker and Hay's (2011) stimuli were composed of standard NZE words that appear in dictionaries.

attention to the association between the words and their age categories, so that listeners were sensitive to the sociolinguistic relationship. The findings in this paper then support that attention driven to stereotypes is an important factor of the social effect in speech recognition (Niedzielski, 1999; Hay et al., 2006a; 2006b; Drager & Kirtley, to appear). Similarly, it is also possible that prominent words in social variation are encoded more strongly by stereotypes, having a high level of cluster robustness, as in the aforementioned Sumner et al.'s (2014) model that posits socially-salient encoding as an independent factor.

What remains uncertain, though, is the timing of the social effect. Since the experiment session was blocked by each speaker, participants heard the two speakers of just one age group during the first half (384 trials), and then the other two speakers for the remaining trials, which was also the case in Walker and Hay (2011). Then, the observed processing advantage could have been drawn by aged-exemplars that are continuously activated prior to the exposure to the stimulus of each trial, rather than activated immediately upon retrieving the signal. For the same reason, the effect is also consistent with speaker adaptation or listeners' expectancy rather than by activation, which is in line with Bayesian models. Even though words were spoken in isolation, the voice itself could have provided more or less contextual information influencing the probability of what the word is. Whether the activation- or probability-based mechanism is responsible for the effect, the independent role of the stereotypical association between words and who produces them should be taken into consideration in either account.

Another question for future work concerns identifying the phonetic features of younger and older voices that trigger the effect. As Walker and Hay (2011) note, the effect could have been caused by a combination of different acoustic properties that cue the ages of speakers. There are, however, some specific features that are controllable using speech synthesis. Not only are there physiological factors, such as generally lower pitch and hoarseness of older females' voices (e.g., Honjo & Isshki, 1980), but also extraneous features involved in age-related variation, such as spectral quality of vowels in Texan English (Gentry, 2006; Pantos, 2006) and in NZE (Gordon & Maclagan, 2001), or laryngeal features of voiceless stops in Seoul Korean (e.g., Kang & Guion, 2008; Kim, 2013; Kang & Han, 2013).

6 Conclusion

The dynamics between the continuously-changing society and human language have drawn considerable interest in sociolinguistics and its related disciplines in cognitive studies. Research in this stream is not only concerned with identifying the social factors and processes of phonetic variation, but also with shedding light on the perceptual mechanism of individual minds that plays a crucial role in the social dynamics. The growing complexity in social interactions raises questions about the cognitive mapping of the dynamics, which is closely related to the nature of speech perception. This study finds that recognition proficiency is improved when a word is produced by speakers whose ages are consistent with the age of the social group who say the words more frequently, and the effect was larger when the speaker age matched the age group that is stereotyped to most likely produce the word. This finding extends our understanding of spoken language perception, providing evidence that lexical representations mirror the social construct in the dynamics. Processing of social information is an important component of word recognition, and lexical access is guided by the cognitive association between social characteristics of words and speakers, including that based on both distribution and stereotypes.

Additional Files

The additional files for this article can be found as follows:

- **Additional file 1: Appendix**. http://dx.doi.org/10.5334/labphon.33.s1

Competing Interests

The author has no competing interests to declare.

References

Bybee, J. L. and Cacoullos, R. T. 2008. Phonological and grammatical variation in exemplar models. *Studies in Hispanic and Lusophone Linguistics, 1,* 399–413. DOI: http://dx.doi.org/10.1515/shll-2008-1026

Cohen-Goldberg, A. M. 2012. Phonological competition within the word: Evidence from the phoneme similarity effect in spoken production. *Journal of Memory and Language, 67,* 184–198. DOI: http://dx.doi.org/10.1016/j.jml.2012.03.007

Creel, S. C. and Tumlin, M. A. 2011. On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65,* 264–285. DOI: http://dx.doi.org/10.1016/j.jml.2011.06.005

Drager, K. 2011. Speaker age and vowel perception. *Language and Speech, 54,* 99–121. DOI: http://dx.doi.org/10.1177/0023830910388017

Drager, K. and Kirtley, M. J. (to appear). Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In: Babel, A. ed., *Awareness and Control*. Cambridge University Press. DOI: http://dx.doi.org/10.1017/cbo9781139680448.003

Ferreira, F. and Clifton, C. 1986. The independence of syntactic processing. *Journal of Memory and Language, 25,* 348–368. DOI: http://dx.doi.org/10.1016/0749-596X(86)90006-9

Foulkes, P. and Docherty, G. 2006. The social life of phonetics and phonology. *Journal of Phonetics, 34*(4), 409–438. DOI: http://dx.doi.org/10.1016/j.wocn.2005.08.002

Gaskell, M. G. and Marslen-Wilson, W. D. 1998. Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 380–396. DOI: http://dx.doi.org/10.1037/0096-1523.24.2.380

Gentry, E. 2006. Hovering between South and West: Houston's merged dialect. Paper presented at New Ways of Analyzing Variation 35. OH: The Ohio State University.

Goldinger, S. D. 1997. Words and voices: Perception and production in an episodic lexicon. In: Johnson, K. and Mullennix, J. eds., *Talker Variability in Speech Processing* (pp. 33–66). San Diego, CA: Academic Press.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. DOI: http://dx.doi.org/10.1037/0033-295X.105.2.251

Gordon, E. and Maclagan, M. 2001. Capturing a sound change: A real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English. *Australian Journal of Linguistics, 21*(2), 215–238. DOI: http://dx.doi.org/10.1080/07268600120080578

Hanna, J. E. and Tanenhaus, M. K. 2004. Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28,* 105–115. DOI: http://dx.doi.org/10.1207/s15516709cog2801_5

Hay, J. and Drager, K. 2010. Stuffed toys and speech perception. *Linguistics, 48*(4), 865–892. DOI: http://dx.doi.org/10.1515/ling.2010.027

Hay, J., Nolan, A. and Drager, K. 2006a. From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 23,* 351–379. DOI: http://dx.doi.org/10.1515/TLR.2006.014

Hay, J., Warren, P., and Drager, K. 2006b. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458–484. DOI: http://dx.doi.org/10.1016/j.wocn.2005.10.001

Honjo, I. and Isshki, N. 1980. Laryngoscopic and voice characteristics of aged persons. *Archives of Otolaryngology, 106*(3), 149–150. DOI: http://dx.doi.org/10.1001/archotol.1980.00790270013003

Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In: Johnson, K. and Mullennix, J. W. eds., *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K., Strand, E., and D'Imperio, M. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27,* 359–384. DOI: http://dx.doi.org/10.1006/jpho.1999.0100

Jun, S-A. 2011. Prosodic markings of complex NP focus, syntax, and the pre-/post-focus string. In: Washburn, M. B., McKinney-Bock, K., Varis, E., Sawyer, A., and Tomaszewicz, B. eds., *Proceedings of the 28th West Coast Conference on Formal Linguistics*, 214–230. Somerville, MA: Cascadilla Press.

Kamide, Y., Altman, G. T. M., Haywood, S. L. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156. DOI: http://dx.doi.org/10.1016/S0749-596X(03)00023-8

Kang, K. and Guion, S. G. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of Acoustic Society of America, 124* (6), 3909–3917. DOI: http://dx.doi.org/10.1121/1.2988292

Kang, Y. and Han, S. 2013. Tonogenesis in early contemporary Seoul Korean: A longitudinal case study. *Lingua, 134,* 62–74. DOI: http://dx.doi.org/10.1016/j.lingua.2013.06.002

Kim, M. 2013. VOT merger and F0 maximization between the lax and aspirated stop in sound change. *The Linguistic Association of Korea Journal, 21*(2), 1–20. DOI: http://dx.doi.org/10.17154/kjal.2013.06.29.2.1

Koops, C., Gentry, E., and Pantos, A. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics, 14,* 93–101.

Kuperman, V., Bertram, R., and Baayen, R. H. 2010. Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language, 62,* 83–97. DOI: http://dx.doi.org/10.1016/j.jml.2009.10.001

Kuperman, V. and Van Dyke, J. A. 2013. Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 802–823. DOI: http://dx.doi.org/10.1037/a0030859

Labov, W. (1972). *Sociolinguistic Patterns*. Pittsburg, PA: University of Pennsylvania Press.

Ladefoged, P. and Broadbent, D. E. 1957. Information conveyed by vowels, *Journal of the Acoustical Society of America, 29,* 98–104. DOI: http://dx.doi.org/10.1121/1.1908694

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. 1989. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*(1), 365–378. DOI: http://dx.doi.org/10.1121/1.397688

Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of language and Psychology*, *18*(1), 62–85. DOI: http://dx.doi.org/10.1177/0261927X99018001005

Norris, D. 2006. The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327–357. DOI: http://dx.doi.org/10.1037/0033-295X.113.2.327

Norris, D. and McQueen, J. M. 2008. Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review, 115*(2), 357–395. DOI: http://dx.doi.org/10.1037/0033-295X.115.2.357

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Phonology: Learning, Memory, and Cognition, 19*(2), 309–328. DOI: http://dx.doi.org/10.1037/0278-7393.19.2.309

Pantos, A. J. 2006. Redefining the South: Teenage Houstonians and the southern shift. Paper presented at New Ways of Analyzing Variation 35, OH: The Ohio State University.

Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. L. and Hopper, P. J. eds., *Frequency effects and the emergence of linguistic structure* (137–158). Amsterdam, Netherland: John Benjamins.

Pierrehumbert, J. B. 2003. Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech, 46,* 115–154. DOI: http://dx.doi.org/10.1177/00238309030460020501

Pinnow, E. and Connine, C. M. 2014. Phonological variant recognition: Representations and rules. *Language and Speech, 57*(1), 42–67. DOI: http://dx.doi.org/10.1177/0023830913479105

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language, 61,* 19–36. DOI: http://dx.doi.org/10.1016/j.jml.2009.02.005

Poushter, J. 2016. Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center*. Retrieved from http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/

Staum-Casasanto, L. 2008. Does social information influence sentence processing? Paper presented at the 30th Annual Meeting of the Cognitive Science Society, Washington, D.C.

Stenovec, T. 2015. The US is way behind much of the world when it comes to internet speeds. *Tech Insider*. Retrieved from http://www.businessinsider.com/akami-global-internet-speeds-rankings-2015-9

Strand, E. and Johnson, K. 1996. Gradient and visual speaker normalization in the perception of fricatives. In: Gibbon, D. ed., *Natural Language Processing and Speech Technology* (pp. 14–26). Berlin, Germany: Mouton de Gruyter. DOI: http://dx.doi.org/10.1515/9783110821895-003

Sumner, M., Kim, S. K., King, E., and McGowan, K. B. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Pscyhology, 4,* 1–13. DOI: http://dx.doi.org/10.3389/fpsyg.2013.01015

Sumner, M. and Samuel, A. G. 2005. Perception and representation of regular variation: The case of final /t/. *Journal of memory and language, 52,* 322–338. DOI: http://dx.doi.org/10.1016/j.jml.2004.11.004

Sumner, M. and Samuel, A. G. 2009. The role of experience in the processing of cross-dialectal variation. *Journal of memory and language, 60,* 487–501. DOI: http://dx.doi.org/10.1016/j.jml.2009.01.001

Tesink, C. M. J. Y., Petersson, K. M., van Berkum, J. J. A., van den Brink, D., Buitelaar, J. K., and Hagoort, P. 2009. Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience, 21,* 2085–2099. DOI: http://dx.doi.org/10.1162/jocn.2008.21161

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580–591. DOI: http://dx.doi.org/10.1162/jocn.2008.20054

Walker, A. and Hay J. 2011. Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology*, *2*(1), 219–237. DOI: http://dx.doi.org/10.1515/labphon.2011.007

Wells, J. 1982. Accents of English. Cambridge, England: Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511611759